

UNCLASSIFIED

AD NUMBER

ADB041699

LIMITATION CHANGES

TO:

Approved for public release; distribution is unlimited.

FROM:

Distribution authorized to U.S. Gov't. agencies only; Administrative/Operational Use; JAN 1971. Other requests shall be referred to Center for Naval Analysis, Arlington, VA.

AUTHORITY

CNA ltr 26 Feb 1980

THIS PAGE IS UNCLASSIFIED

FOR OFFICIAL USE ONLY

02 000167.00

COMPUTER PROGRAMS FOR ANALYZING MANPOWER DATA

Peter H. Stoloff

Research Contribution 167

***Center
for
Naval
Analyses***

Institute of Naval Studies

an affiliate of the University of Rochester

1401 Wilson Boulevard, Arlington, Virginia 22209 — Contract N00014-68-A-0091

This Research Contribution does not necessarily represent the opinion of the Department of the Navy.

FOR OFFICIAL USE ONLY

1401 Wilson Boulevard

Arlington, Virginia 22209

703/524-9400

**Center
for
Naval
Analyses**

*an affiliate of the
University of Rochester*

(INS)89-71
23 March 1971

To: Distribution List

Subj: CNA Research Contribution #167, "Computer Programs for
Analyzing Manpower Data" by Peter Stoloff

Encl: (1) One (1) copy of subject RC

1. Enclosure (1) is forwarded as a matter of possible interest.
2. This Research Contribution was developed in connection with the Institute of Naval Studies Manpower Studies program. The program file presented here consists of computer programs developed to facilitate machine extraction and processing of Five Year Defense Program and other manpower data.
3. Research Contributions are distributed for their potential value in other studies or analyses. They have not been subjected to extensive internal CNA review and do not represent the opinion of the Department of the Navy.



DAVID KASSING
Director
Institute of Naval Studies

(INS)89-71
23 March 1971

Subj: CNA Research Contribution #167, "Computer Programs for
Analyzing Manpower Data" by Peter Stoloff

DISTRIBUTION LIST:

DIR, ARPA
COMDT AFSC
ADMIN, DDC (2)
COMDT NATL WAR COLLEGE
Office of Science and Technology (EXOP)
National Science Foundation (EXOP)
Federal Aviation Agency (Dept of Transportation)
Scientific & Technical Information Facility, NASA

Office of the Comptroller (Dept of the Navy)
Ass't SecNav (R&D)

OPA

ONR

Op-00	Op-01
Op-09	Op-03
Op-09B9	Op-91
Op-09M	Op-090
Op-90	Op-503
Op-905	Op-098T
Op-96	

Fleet Computer Programming Center Atlantic
Fleet Computer Programming Center Pacific
Navy Medical Research Institute
Navy Aerospace Medical Center
Navy Personnel Programs Support Activity
Navy Personnel Training Research Laboratory
U. S. Naval Academy
Naval Postgraduate School
Naval War College
Commandant, U. S. Coast Guard
Department of the Army (Attn: Adj Gen'l) (6)
Commandant, Army War College
Department of the Air Force (Attn: PDO 4008) (9)
Superintendent, Air University
Center for Research in Social Systems (CRESS)
Human Resources Research Office (HUMRRO)
Institute for Defense Analyses (IDA)
Research Analysis Corporation (RAC)
The RAND Corporation
Stanford Research Institute (SRI)
Systems Development Corporation (SDC)

CENTER FOR NAVAL ANALYSES

Institute of Naval Studies

RESEARCH CONTRIBUTION 167

**Computer Programs for
Analyzing Manpower Data**

Peter H. Stoloff

January 1971

Work conducted under contract N00014-68-A-0091

Enclosure (1) to (INS)89-71 dated 23 March 1971

ESTIMATION OF NAVY SUPPORT MANPOWER REQUIREMENTS
(SOMAR)

Robert Lockman, Project Director
Robert K. Lehto, Op-96 Study Monitor

ABSTRACT

The MANPOWER program file consists of computer programs developed to facilitate machine extraction and processing of Five-Year Defense Program and other manpower data. The file includes routines for data file building, manipulation, and data transformation and analysis.

A set of utility routines is described which allows the user to extract and execute any program on the MANPOWER file without having to handle any of the source program card-decks. Using these routines, the MANPOWER program file, and a data file, the entire process of building and analyzing a data bank could be accomplished as a single job. This system has been developed for use on Control Data 3400, 3600, and 3800 machines. Modifications to some of the programs might be necessary if other equipment were used.

TABLE OF CONTENTS

	Page
Summary of programs	1
Program PESORT	3
General description	3
Output	3
Limitations	3
Input requirements	3
Card order	3
Specific card preparation	4
Program PE	9
General description	9
Output	9
Limitations	9
Input requirements	9
Card order	9
Specific card preparation	10
Program MOMENTS	13
General description	13
Output	13
Limitations	13
Special features	13
Card order	13
Specific card preparation	14
Computational procedure	17
Central moments	17
Correlation and regression analyses	18
Transformation codes	19
Program CLUSTER	21
General description	21
Output	21
Limitations	21
Special features	21
Card order	22
Specific card preparation	22
Computational procedure	23
Program PEXGROUP	27
General description	27
Output	27
Input requirements	27
Limitations	27

TABLE OF CONTENTS (Cont'd)

	Page
Card order	27
Specific card preparation	28
Computational procedure	28
Program REGFIT	31
General description	31
Output	31
Input requirements	31
Specific card preparation	31
Computational procedure	31
Program ITERFIT	41
General description	41
Output	41
Input requirements	41
Limitations	41
Card order	42
Specific card preparation	42
Computational procedure	44
Manpower utility routine	47
Program INTERFAC	49
General description	49
Input/output	49
Limitations	49
Card order	49
Specific card preparation	49
Algorithm	50
Output phase	51
Program ONTAPE	53
General description	53
Output	53
Input	53
References	55

SUMMARY OF PROGRAMS

These computer programs were written in support of INS Study 31, "Estimation of Navy Support Manpower Requirements (C)", (reference (d)). Although some of the programs were developed specifically for building and processing the study data, many are of a generally useful nature and probably could be adapted easily to almost any computer. The data analysis routines accept either card images or formatted tape input, making them available for the user without modification. A short summary of each program contained in the package, referred to as the MANPOWER PROGRAM FILE, follows.

PESORT and PE are file building and editing routines used for processing raw FYDP data into a form usable for analytic purposes.

MOMENTS is a generalized program and provides for computing central moments, intercorrelations, and multiple regression equations for a data matrix with as many as 100 variables and unlimited observations. It can be used to plot and print distributions and univariate histograms, as well as bivariate data plots.

CLUSTER is a generally useful program for the grouping and reduction of data. Algorithms for grouping elements similar to one another and dissimilar to other elements are described. Various indexes of similarity of data elements may be used in the grouping process.

PEXGROUP, although written for the purpose of computing comparative statistics between FYDP program elements and composites of program elements, can be easily modified to compute statistics including part-whole correlation coefficients and other measures used in test construction research.

REGFIT and ITERFIT are both nonlinear regression programs. Iterative routines are used to determine least squares parameters for fitting a set of data points and for extrapolating beyond the range of the data.

The utility routines are to facilitate card and tape handling for both the FYDP analyst or general user of these computer programs.

PROGRAM PESORT

GENERAL DESCRIPTION

This program processes "raw" FYDP data by scanning the file for logical inconsistencies, and performs a preliminary sort of the data.*

Output

- Listing of the program element (PE) dictionary;
- Listing of the preliminary sort categories;
- Tally of number of PE's per sort category;
- Summary of data-by service (Navy, Marine Corps);
- Summary by first digit of PE code;
- Magnetic tape file of the processed FYDP data.

Limitations

- Processes each year sequentially for a given group of PE's.

INPUT REQUIREMENTS

- Dictionary. A dictionary of FYDP PE's containing the PE numbers and their corresponding mnemonics.

- Group Identifiers punched with a sort-category label. Cards interspersed with the PE dictionary deck are used to denote the beginning of a category grouping of PE's.

- Data file. The card images, loaded onto magnetic tape, represent manpower counts and Total Obligational Authority (TOA) values for a single PE for a single year. The hierarchical ordering of FY by PE is required, i.e., all PE's for a given fiscal year are adjacent records on the data file. A special code, -1000, in the PE identification fields serves as an end-of-year marker. LUN I is designated as the data input file.

CARD ORDER

Set

1. Descriptive information
2. Parameter specification
3. Number of PE's/category
4. Group names, PE dictionary
5. FYDP DATA (on tape)

*This program has the capability of producing a final sort, but other programs described later do this more efficiently. The logical consistency checks are the primary functions performed by this program.

SPECIFIC CARD PREPARATION

<u>Sequence</u>	<u>Contents/format</u>
1.	Descriptive information <ul style="list-style-type: none"> Any alphanumeric punches in columns 1-80.
2.	Parameter specification
<u>Columns</u>	<u>Contents</u>
1-3	Number of categories (NC)
4-6	Blank
7-9	Number of years of data on input file
10-12	Any 3-digit code used as identification on output file
13-15	Code controlling output file <p>0 punch = no tape copy of summary (by category) data > 2 punch = tape copy will be produced on LUN 2</p>
16-18	Number of files to be skipped on input file
19-21	PUNCH 2 if columns 13-15 > 1
23-25	A blank or zero punch suppresses output tape when input (FYDP) file is sorted by PE's within categories. Punch 1 if this output is needed. Output will be on LUN 2.
25-27	Year corresponding to starting date of input data file, e.g., 62 = 1962.
3.	Number of PE's per category. A list with the number of elements corresponding to the value punched in columns 1-3 of the Parameter specification card. $\left[n_i \mid i = 1, NC \right]^*$ Format: (25I3)
4.	Group names (labels) cards interspersed with PE dictionary. A label card, followed by $\left[n_i \mid i = 1, NC \right]$ PE dictionary cards, is prepared as follows: Label: Punch any alphanumerics in columns 1-80 which will serve as a label for the PE grouping or category. Follow this by the set of n_i cards containing the PE name in columns 1-28, with identification code in columns 29-34.
5.	FYDP data file. The current version of PESORT requires that the FYDP data be formatted as follows:

*This notation is used to indicate a number, n_i , which is the ith element of a set of NC values of n.

<u>Column/position</u>	<u>Contents</u>
1-2	Fiscal year (e.g., 1969 would be punched as 69).
5-10	PE 6-digit code
11-20	TOA (non-discounted)
21-28	Navy - number officers
29-36	Navy - number enlisted
37-44	Marines - number officers
45-52	Marines - number enlisted
53-60	Civilians - direct hire - U.S.
61-68	Civilians - direct hire - foreign
69-76	Civilians - contract
77-80	Blank

These are card images loaded on 7-channel, 556-BPI, BCD magnetic tape as unblocked (single) records. The data are "stacked" so that all PE's for a given year are contiguous on the tape. A card punched with -1000 starting in column/position 6 is used as an end-of-year marker. The last end-of-year marker on the tape is followed by an end-of-file marker.

SAMPLE FORMAT FOR HYPOTHETICAL FYDP DATA

Card number

Layout

cc 1-41

1

Test Δ case, Δ 5 Δ PE's, Δ 2 Δ categories, Δ 2 Δ years.

cc	3	6	9	11	12	15
	2	5	2	6	8	1

2

cc	3	6
	2	3

3

cc	1-7
	Group Δ I

4

cc	1-28	29-34
	PE-A	000011
	PE-B	000021

5

6

cc	1-8
	Group Δ II

7

cc	1-28	29-34
	PE-C	000031
	PE-D	000041
	PE-E	000051

8

9

10

cc/ tp	1-2	Variable
	68	000011
	68	000021
	68	000031
	68	000041
	68	000051
		-1000
	69	000011
	69	000021
	69	000031
	69	000041
	69	000051
		-1000

11

12

13

14

15

16

17

18

19

20

21

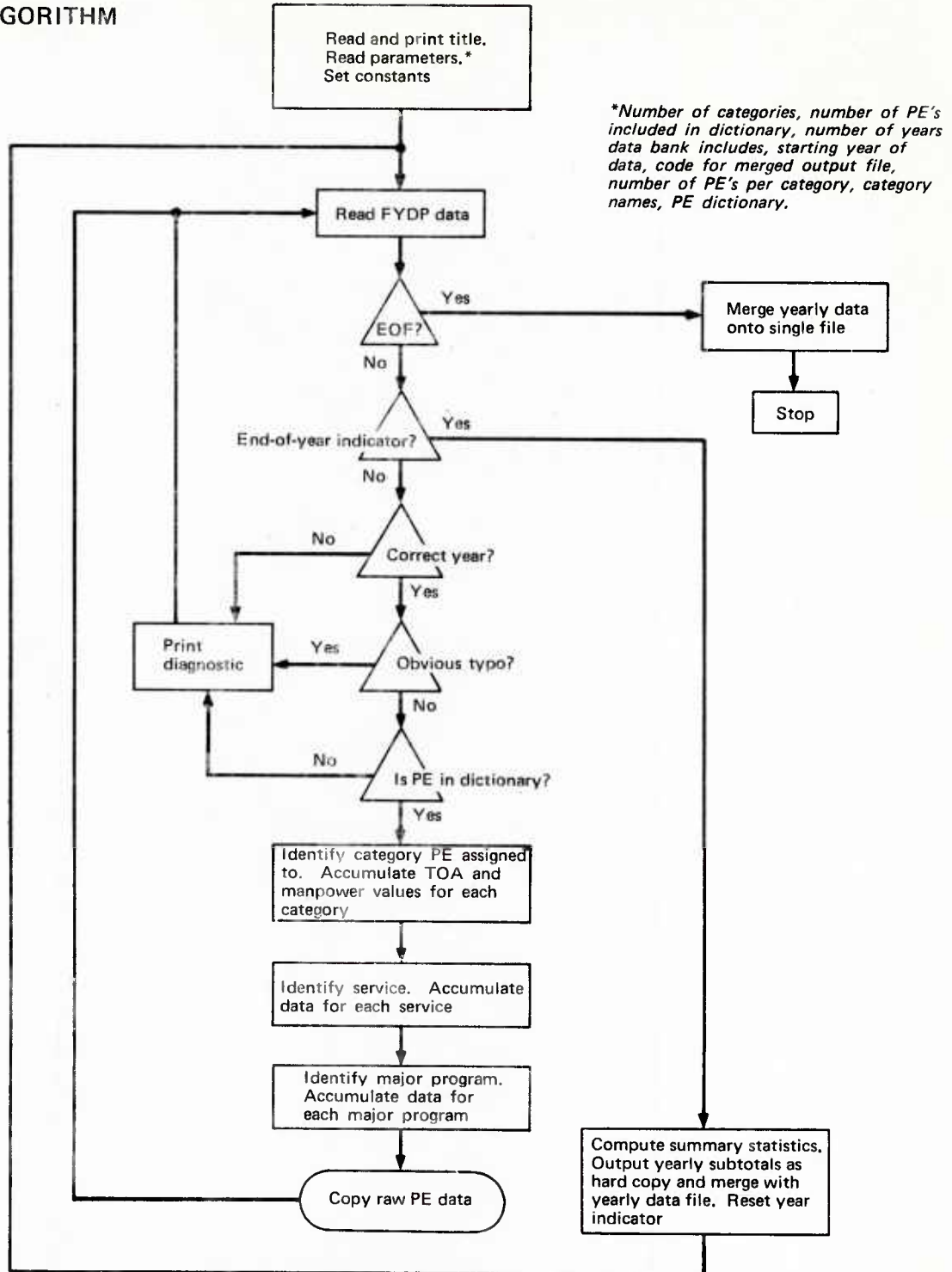
22

23

EOF

NOTE: cc denotes card column; tp, tape position.

ALGORITHM



PROGRAM PE

GENERAL DESCRIPTION

An extension of Program PESORT, PE is the key program used to process the raw FYDP data as input to the analytical routines. The data processing consists of 3 phases: categorization/sorting, merging, and display. PE groupings or categories are defined by the user, and FYDP data for each category are aggregated into yearly composites as hard copy (printout).

Output

- Listing of PE dictionary, elements grouped by categories;
- Magnetic tape file of the FYDP data;
- Hard copy display of data, summarized by PE category (see table 1);
- Merged FYDP (by category) tape-file.

Limitations

- A maximum of 450 PE's. The sum of PE categories and composites of the categories may not exceed 50.
- A maximum of 15 years of FYDP data may be processed with a single pass of the input file.

INPUT REQUIREMENTS

- Dictionary. (See PROGRAM PESORT)
- Data file. (See PROGRAM PESORT.) Mount tape on LUN 1.
- Group identifiers. (See PROGRAM PESORT.)
- Composite request cards. Composites consisting of the sums of elements of different categories are obtained by indicating the indexes of the categories to be included in a composite on this card.

CARD ORDER

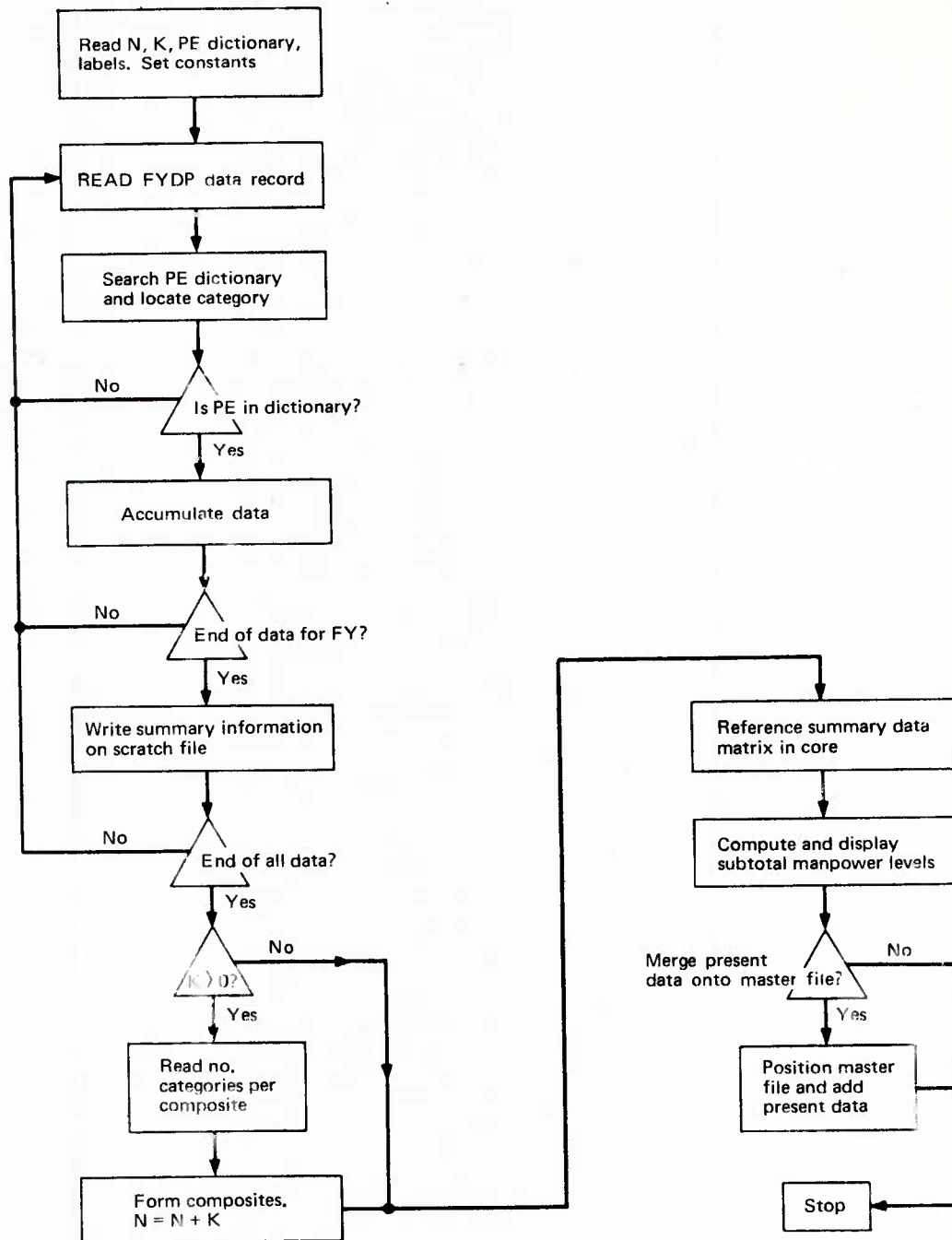
Set

1. Parameter specification
2. Number PE/category
3. Category group names, PE dictionary
4. Indexes of categories for each composite
5. Names for components

SPECIFIC CARD PREPARATION

<u>Sequence</u>	<u>Contents/format</u>																
1.	Parameter specification																
	<table> <tr> <th><u>Columns</u></th><th><u>Contents</u></th></tr> <tr> <td>1-3</td><td>Number of categories (N), (≤ 50)</td></tr> <tr> <td>4-6</td><td>Number of composites to be formed (K), (≤ 25)</td></tr> <tr> <td>7-9</td><td>Blank</td></tr> <tr> <td>10-12</td><td>Number of FY's represented on input data file</td></tr> <tr> <td>13-15</td><td>Code representing FY of earliest FYDP data on the</td></tr> <tr> <td>16-18</td><td>Logical unit number (LUN) containing pre-processed FYDP data (by category) to be merged with output of this job.</td></tr> <tr> <td>21</td><td>To eliminate list of PE dictionary, punch an integer ≥ 1.</td></tr> </table>	<u>Columns</u>	<u>Contents</u>	1-3	Number of categories (N), (≤ 50)	4-6	Number of composites to be formed (K), (≤ 25)	7-9	Blank	10-12	Number of FY's represented on input data file	13-15	Code representing FY of earliest FYDP data on the	16-18	Logical unit number (LUN) containing pre-processed FYDP data (by category) to be merged with output of this job.	21	To eliminate list of PE dictionary, punch an integer ≥ 1 .
<u>Columns</u>	<u>Contents</u>																
1-3	Number of categories (N), (≤ 50)																
4-6	Number of composites to be formed (K), (≤ 25)																
7-9	Blank																
10-12	Number of FY's represented on input data file																
13-15	Code representing FY of earliest FYDP data on the																
16-18	Logical unit number (LUN) containing pre-processed FYDP data (by category) to be merged with output of this job.																
21	To eliminate list of PE dictionary, punch an integer ≥ 1 .																
2.	Number of PE's per category. Punch, right justified, in fields of 3 columns, a maximum of 25 per card. Format: (25I3). Continue on another card if necessary.																
3.	Category, group names, PE dictionary. (See PROGRAM PESORT for further details.)																
4.	} Prepare these cards as described in PROGRAM PESORT.																
5.																	

ALGORITHM



PROGRAM MOMENTS

GENERAL DESCRIPTION

This program provides a comprehensive statistical and graphical description of a set of up to 100 variables for as many as 9999 observations.

Output

- Number of observations having complete data;
- First 4 central moments, standard error, t values for skewness and kurtosis for each variable;
- Zero-order intercorrelations among variables;
- Multiple regression equations for prespecified group of variables;
- Scatter plots of pairs of variables, with or without identification of individual data points;
- Histograms (bar graphs) of univariate distributions.

Limitations

- Total number of variables before and after transgeneration (k) ≤ 100 ;
- Total number of observations per variable (N) ≤ 9999 ;
- If plots or histograms are desired, $N \cdot K \leq 10000$;
- Number of variable format cards ≤ 9 .

Special Features

- Scans for missing data (blanks are missing data). Drops an observation from sample statistics if one or more variables is "blank;"
- Allows for transformation of variables (e.g., log, square root);
- Variables may be labeled with a 6-character alphanumeric mnemonic.

CARD ORDER

1. Descriptive information card
2. Parameter specification card
- *3. Variable name card or cards
4. Variable format card or cards
- *5. Transformation card or cards
- *6. Data cards (if card input)
- *7. REGRES card or cards
- *8. PLOT selection card or cards
9. FINISH card

(Repeat 1 through 8 for additional subanalyses.)

SPECIFIC CARD PREPARATION

Sequence

Type

1. Descriptive information card
 - Any alphanumeric characters in columns 1-72
2. Parameter specification card

<u>Columns</u>	<u>Parameter</u>
1-4	Number of observations
5-7	Number of variables before transformation
8	Number of variable format cards
9-11	Number of plots and histograms
12-13	Logical unit number of data input (other than 1)
15	Rewind input unit when punched 0 or blank
16-17	Number of transformation cards
18-19	Number of variables after transformation
21	To abort correlation & regression analysis, punch 1
23	To read names for variables, punch 1
24-25	Number of regression analyses

Sequence

Type

- *3 Names for variables - (see column 23 on Parameter specification) card

<u>Columns</u>	<u>Parameter</u>
1-6	Alphanumeric code for 1st variable
7-12	Alphanumeric code for 2nd variable
67-72	Alphanumeric code for 12th variable
--	Use additional cards to label all variables retained for analysis after transformation. (columns 18-19 on Parameter specification card)

4. Variable format card or cards

Punch format of data in columns 1-72 (column 1 of first card must contain an open parenthesis)

- *5 Transformation card or cards (see columns 16-17 on Parameter specification card)

<u>Columns</u>	<u>Parameter</u>
1-5	TRANS
7-9	Code, indicating type of transformation--see page B-9
10-12	Index number of the input variable
13-15	Index number of the output variable
16-25	A constant--if relevant to the transformation (punch decimal point, if applicable)

- *6 Data cards--if data not on tape

<u>Sequence</u>	<u>Type</u>
-----------------	-------------

- *7 Regression analysis card or cards (if column 25 \geq on Parameter specification card) **

<u>Columns</u>	<u>Parameter</u>
1-6	REGRES
7-9	Index number of the criterion (dependent) variable
10-12	Total number of predictor (independent) variables
13-15	Index number of 1st predictor variable
16-18	Index number of 2nd predictor variable
70-72	Index number of 20th predictor variable

After the last regression analysis desired, insert the following card:

<u>Columns</u>	<u>Parameter</u>
1-4	ENDR

8. PLOT selection card or cards

<u>Columns</u>	<u>Parameter</u>
1-4	PLOT
7-9	Index number of the variable as ABSCISSA, X
10-12	Index number of the variable as ORDINATE, Y
13-15	YES if points are to be identified
16-18	YES if page divider is to be printed

(Note: if X = Y, the program builds a histogram. If a bar graph is desired, punch YES in columns 13-15)

After last PLOT selection card, include the following card:

<u>Columns</u>	<u>Parameter</u>
1-4	ENDP

9. FINISH card. More than one analysis may be processed as a single job. Repeat cards 1 through 8 for additional subproblems. After the last card of the last analysis, the following card should be included:

1-6	FINISH
-----	--------

* Optional depending upon contents of Parameter specification card.

** For more than 20 predictors, prepare as many REGRES cards as needed. For exactly 20, or multiples of 20 predictor variables, follow the REGRES cards with a blank card.

COMPUTATIONAL PROCEDURE

Central Moments

Let:

N = Number of observations

K = Number of variables

X_{ij} = ith observation on jth variable

(The summation operator is across N observations and the subscript denoting the jth variable is dropped.)

$$S = \sum X$$

$$S2 = \sum X^2$$

$$S3 = \sum X^3$$

$$S4 = \sum X^4$$

The mean (\bar{X}) of a given variable is computed as S/N ; and the standard deviation (SD) as $\left[\frac{(S2 - S^2/N)}{N - 1} \right]^{1/2}$

Let:

$$K3 = S3 - 3(S2/N) + 2(S3/N^2)$$

$$K4 = S4 - 4 \cdot S(S3/N) + 6(S^2 \cdot S2/N^2) - 3(S^4/N^3)$$

Skewness (G1) is computed as $K3/(SD \cdot SD^2)$

Kurtosis (G2) is computed as

$$\left\{ \frac{N[(N+1) \cdot K4 - 3(N-1)(S2 - S^2/N)^2]}{(N-1)(N-2)(N-3)} \right\} \div SD^3$$

The standard error (SE) of G1 is $[6N(N-1)/(N-2)(N+1)(N+3)]^{1/2}$

The standard error of G2 is $[24N(N-1)^2/(N-3)(N-2)(N+3)(N+5)]^{1/2}$

Values for t, at d.f. equal to ∞ , are given by

$$t_{G1} = G1/SE_{G1}, \text{ and}$$

$$t_{G2} = G2/SE_{G2}$$

Correlation and Regression Analyses

Zero-order intercorrelations among the ij th pair of variables for all pairs are computed by:

$$r_{ij} = \frac{N \sum_{i=1}^n \sum_{j=1}^n X_i X_j - \sum_{i=1}^n X_i \sum_{j=1}^n X_j}{\sqrt{N \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2} \cdot \sqrt{N \sum_{j=1}^n X_j^2 - (\sum_{j=1}^n X_j)^2}}$$

Multiple regression equations are constructed as follows:

- Intercorrelations among the M independent variables (R) are selected from the matrix containing intercorrelations among all variables.
- A vector, v , of correlations of the dependent (d) and M independent variables is also selected.
- The vector of standardized regression weights, β , is computed by the relationship

$$\beta = R^{-1}v$$

- Raw score regression weights are computed by

$$b_j = \frac{\sigma_d}{\sigma_j} \beta_j$$

- The regression intercept, C , is determined by

$$C = \bar{X}_d - \sum_{j=1}^M b_j \bar{X}_j$$

- The multiple correlation coefficient $r_{d \cdot M}$ is given by

$$r_{d \cdot M} = \left(\sum_{j=1}^M \beta_j v_j \right)^{1/2}$$

- An F test for the significance of $r_{d \cdot M}$ is given by

$$F = \frac{r_{d \cdot M}^2 (N-M)}{(1 - r_{d \cdot M}^2) (M-1)}$$

which has $(M - 1)$, $(N - M)$ degrees of freedom and $d \cdot M$ denotes the correlation of the M independent variables with the dependent variable, d .

- The relative contribution of each variable to the predictable

variance $r_{d \cdot M}^2$, is given for reference.

TRANSFORMATION CODES

SUBSCRIPT "OUT" indicates index number corresponding to where the transformed variable is stored; SUBSCRIPT "IN" indicates the index number of the input variable which is to be operated upon. C is a numerical constant.

<u>Code</u>	<u>Resulting Transformation</u>
1	$X(OUT) = SORT(X(IN))$
2	$X(OUT) = ALOG(X(IN))$
3	$X(OUT) = 1.0/X(IN)$
4	$X(OUT) = SORT(X(IN) + C)$
5	$X(OUT) = ALOG(X(IN) + C)$
6	$X(OUT) = X(IN)^C$
7	$X(OUT) = ARCSIN(X(IN))$
8*	$X(OUT) = X(OUT) * X(IN)$
9*	$X(OUT) = X(OUT) * X(IN) * C$
10*	$X(OUT) = X(OUT) + X(IN)$
11	$X(OUT) = X(IN) + C$
12	$X(OUT) = 1.0/(X(IN) + C)$
13	$X(OUT) = EXP(X(IN))$

*Assumes X (OUT) was created by a previous transformation.

PROGRAM CLUSTER

GENERAL DESCRIPTION

This program performs a cluster or profile analysis on a set of measurements for a given sample of individuals or units to identify any mutually exclusive subgroups of either sampling units or variables measured.

Output

- Means and standard deviations of profiles and variables;
- Standard scores for profile elements/variables;
- Generalized distances between elements/variables;
- Congruency coefficients between elements/variables;
- Cattell's r_p coefficients between elements/variables;
- Covariances between elements/variables;
- Correlations between elements/variables;
- Clusters based upon
 - correlations
 - congruency coefficients
 - r_p
- Composite scores based upon cluster membership;
- Mean correlations of profiles with clusters;
- Mean correlations between clusters.

Limitations

- The number of variable format cards ≤ 9 ;
- The product of sample size (N) and the number of elements per profile (K) $\leq 9,000$.

Special Features

- Scans for missing data (blanks) and deletes a score profile if one or more elements are missing.
- Standardizes score profile if requested.

CARD ORDER

1. Descriptive information card
2. Parameter specification card
3. Variable format card or card
4. Data
5. FINISH card

(Repeat 1 through 4 for additional subanalyses. Unit 5 should be "equipped" to 60, and unit 6 to 61.)

SPECIFIC CARD PREPARATION

Sequence

Type

1. Descriptive information card
 - Any alphanumeric data in columns 1-72.
2. Parameter specification card

Columns

Parameter

1-3	Number of sampling units (N)
4-6	Number variables (K)
9	2 punch = calculate and print D^2
12	1 punch = calculate and print congruency coefficients
	2 punch = print and use congruency coefficients as basis for clustering
15	1 punch = print covariance matrix
18	2 punch = calculate, print and use Pearson r for clustering
19-21	Alternate tape unit (T) for data input ($50 < T < 6$)
22-26	C(L) - lower limit on r for including a profile in a cluster
27-31	C(U) - upper limit on r for excluding a profile from a cluster $[C(L) > C(U)]$
36	1 punch = do not rewind alternate input tape
39	1 punch = print standard score profiles
*41	1 punch = punched output
42-43	Number of variable format cards

*Not recommended.

45	0 punch = r_p as Cattell profile similarity coefficient
46-53	Chi Square value for determining r_p (when omitted, coefficient is not printed)
54-57	C(L) value for r_p
58-61	C(U) value for r_p
62-63	2 punch = transpose data matrix (cluster by variables, rather than by score profiles)
65	1 punch = profiles are <u>not</u> to be standardized
67	1 punch = uses absolute value of r as basis of clustering
68-69	Minimum number of profiles needed to form a cluster (if blank, program sets it equal to 3.)

<u>Sequence</u>	<u>Type</u>
-----------------	-------------

3. Variable format card or cards
4. Data (cards or tape)
5. FINISH card

- Last card in deck must contain:

<u>Columns</u>	<u>Punch</u>
1-6	FINISH

COMPUTATIONAL PROCEDURE

Let X be a matrix of scores, having n rows representing the number of subjects or sampling units, and k columns representing the number of variables comprising the score profile for a given sampling unit. A given column of X may be represented as a vector, x_j , where $j = 1, 2, \dots, n$, which contains n sampling units for the j th variable.

The standard score profile Z_i , representing the i th row or profile of X , is given as

$$Z_i = [x_i - \text{Mean}(x)] / \sigma(x)$$

This standardization procedure is equivalent to rescaling all elements in the i th profile, such that the mean, or elevation, of the profile is zero and its variance is unity.

At the discretion of the user, profile standardization may be omitted. (If one is clustering variables, rather than sampling units, standardization is the exception, rather than the rule.) A number of different similarity coefficients are computed. A generalized distance function between the l th and m th elements in X (by either rows or columns, depending upon the option elected), is computed by

$$D_{lm}^2 = \sum_i^n x_{il}^2 + \sum_i^n x_{im}^2 - 2 \left(\sum_i^n x_{il} x_{im} \right) / n.$$

The congruency coefficient, a measure of profile similarity, is computed as

$$G_{lm} = \sum_i^n (z_{il} z_{im}) / \sqrt{\sum_i^n z_{il}^2 \times \sum_i^n z_{im}^2},$$

where Z are scores standardized across sampling units.

Covariances and Pearson product moment correlations are computed by the usual method.

The final similarity coefficient which may be computed is Cattell's r_p (reference (a)):

$$r_p = \frac{E_k - \sum Z_{lm}^2}{E_k + \sum Z_{lm}^2},$$

where E_k is twice the median Chi Square value (i.e. $p = .5$) for k degrees of freedom ($E_k \approx 2k$). (A table showing the median Chi Square value for different values of k , and the probability of exceeding r_p , is given in reference (a).)

The present illustration uses the Pearson product moment correlation coefficient as the similarity measure between profiles as the basis for clustering.

Search for an initial cluster begins by selection of a profile that correlates at or above the value $C(L)$. The limit of $C(L)$ may be set at the value at which a correlation coefficient based upon k independent variates is significant at level α . The profile having the greatest number of significant correlations with other profiles is selected as a pivot. To the pivot is added the profile with the highest average correlation with all members in the pivot list. To this first pair is added the profile that correlates highest on the average with members of the list belonging to the pair. The entire remaining $(n-2) \times (n-2)$ correlation matrix is searched for the profile that has the highest average correlations with those already in the cluster and this profile is added to the cluster. This procedure is repeated until there are no other profiles whose average correlations with members of the cluster equals or exceeds $C(L)$.

Next, an upper limit $C(U)$, which is a cutting point used to define dissimilarity, is set to prevent cluster overlap. A suitable value is the point where a correlation coefficient is significant at some p value, $> p$ for $C(L)$. Any profile in the residual matrix, i.e., those not already selected for the J -1th cluster, that correlates on the average $\geq C(U)$, is deleted and is not considered for any subsequent clusters. The clustering process continues until at least m members can be formed (set by the user, where $m \geq 3$). (See Lorr (reference (b)) for applications of this method.)

PROGRAM PEXGROUP

GENERAL DESCRIPTION

This program computes part-whole correlations between a composite (category) and each item (PE) in the composite.

Output

- $r_{i,c}$ (i = item; c = composite) - item, composite correlation
- $r_{c,i}$ - correlation of item and composite with the variance in c, due to i held constant;
- List of the data for each PE in the FYDP data bank;
- Means and standard deviations;
- Count of data points for each PE.

Input Requirements

- FYDP data file, as described in PROGRAM PESORT. Mount tape on LUN 1.
- PE dictionary, as described in PROGRAM PESORT.

Limitations

- Will only perform an analysis of a single variable or a single composite for a single pass of the data file.

CARD ORDER*

1. Descriptive information card
2. Parameter specification card
3. Group category size card
4. Group name cards
PE-dictionary
5. FINISH

* Repeat card (sets) 1-2 only for analysis of different composites of variables from data file.

SPECIFIC CARD PREPARATION

<u>Sequence</u>	<u>Contents/format</u>
1.	Descriptive information <ul style="list-style-type: none"> • Any alphanumeric characters in columns 1-80
2.	Parameter specification card
Columns	Contents
1-3	Number of categories
4-6	Number of PE's included in data file
5-9	Number of composites to be formed
10-12	Index number of the 1st variable in the composite
13-15	Index number of the 2nd variable in the composite (in fields of 3 columns for up to 8 variables)
3.	Group size card (2513). The number of PE's comprising each category
4.	Group names - PE dictionary. See PROGRAM PESORT for detailed information.
5.	FINISH card - punch - FINISH in columns 1-6 to indicate end-of-job

COMPUTATIONAL PROCEDURE

Let C represent a composite of \underline{n} items designated,

$$\left\{ x_i \mid i = 1, M \right\} \text{ and } C = \sum_{i=1}^n x_i$$

The correlation of an x_i with C is given by the product moment correlation,

$$r_{c \cdot x_i}$$

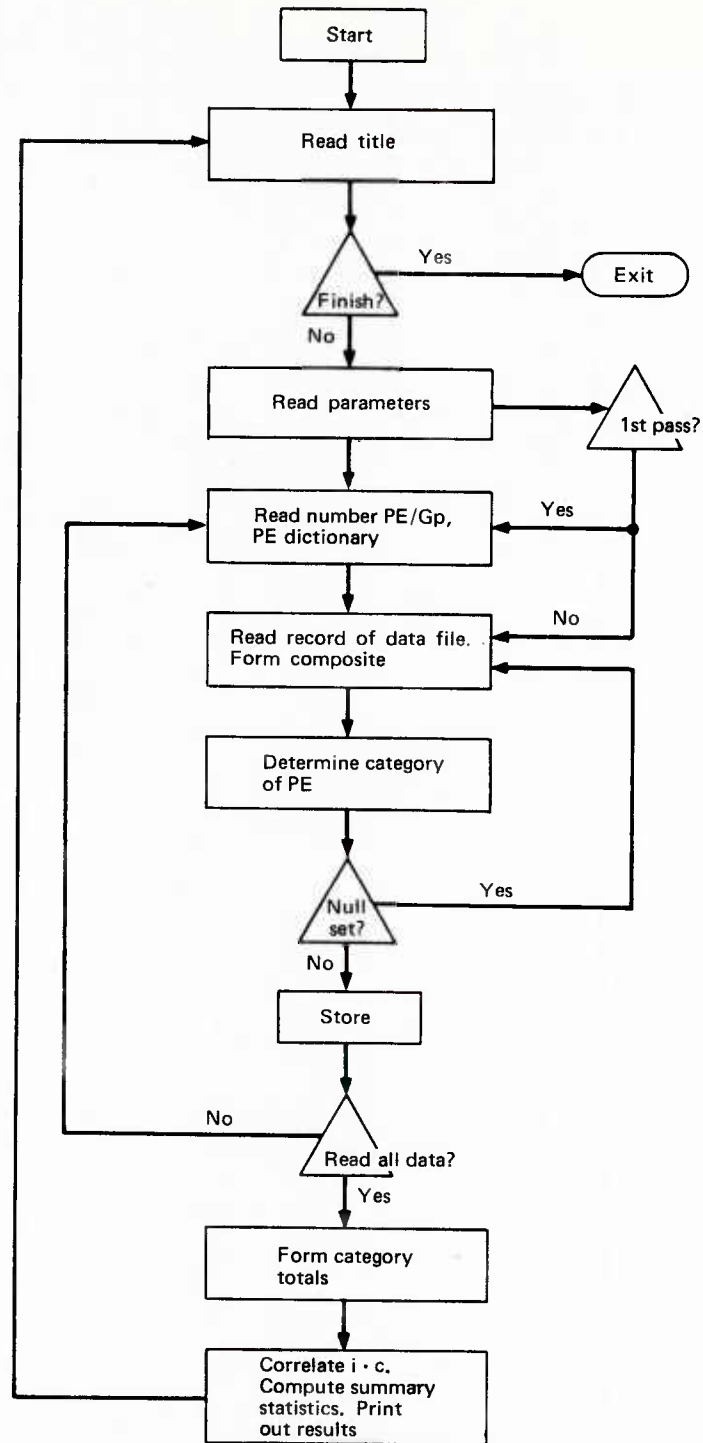
The correlation of x_i and C, independent of the variance of x_i is determined by

$$\frac{r_{c \cdot x_i} = r_i, \quad c \sigma_{x_i} - \sigma_c}{\sqrt{\sigma_{x_i}^2 + \sigma_c^2 - 2r_{i,c} \sigma_{x_i} \sigma_c}},$$

where

$$x_i \in x_c.$$

ALGORITHM



PROGRAM REGFIT

GENERAL DESCRIPTION

REGFIT is a generalized routine for computing the coefficients for an m^{th} order polynomial using a least squares regression method. The routine computes polynomials for the entire range of 1 through m and evaluates goodness of fit after each iteration.

Output

- Means and standard deviations of Y_n (the dependent variable) and X^1 through X^m (powers of the independent variable);
- Raw (b) and standardized (B) regression weights;
- Product-moment correlation between X and \hat{Y} (predicted or fitted dependent variable);
- Y intercept, standard error of estimate;
- Various tests for goodness of fit;
- \hat{Y} values and confidence limits for user specified values of X , not necessarily within the range of the original X values;
- Optional listing of X , Y , and Y -predicted values.

Input Requirements

- X and Y data points (variable format).

Limitations

- The array containing original X , Y values is limited to 50 variables and 20 observations per variable.
- $m \leq 10$
- Missing data (observations) are treated as real, zero values.

SPECIFIC CARD PREPARATION

Sequence

Contents/format

- | | |
|----|--------------------------------------------------------|
| 1. | Title card--any descriptive information (columns 1-80) |
| 2. | Parameter specification |

<u>Columns</u>	<u>Contents</u>
1-3	N (number observations)
4-6	M (number variables)
7-9	Input unit number of data file (blank = cards)
11-12	FY of earliest data contained on input file; (e.g., 63 = 1963).
13-15	Blank or zero, produces estimates rounded to nearest unit. Punch ≥ 1 , output in E 20.8 format
3.	Variable format card (columns 1-80)
4.	DATA (if on cards)
5.	FIT cards
<u>Columns</u>	<u>Contents</u>
1-6	Any non-blank characters
*7-9	Index of X (independent) variable
10-12	Index of Y (dependent) variable
13-15	M-degree of highest order polynomial desired
16-17	Punch NO if listing of X, Y and Y-predicted not wanted
18-19	Any non-blank characters to add additional data point; X = Y = 0
20-21	A code to indicate status of prediction option.
<u>Code</u>	<u>Prediction option**</u>
0	No prediction
1	List of X values are to be read
2	A range of X values are used.

* Punch zero (0) if X is a time series.

** See 5A, 5B, and 5C.

Sequence

5.

22-23

Code (K) specifying the subset of observations selected from data bank. When $K \leq 0$, all observations are used, when $K \geq 2$, subsets of data are selected. See 5D.

24-25

A code controlling computation of the standard error of Y predicted from X.

Code

Result

0, blank $SE^2 = \sigma^2 Y(1 - R^2)$

≥ 1 $SE^2 = \sigma^2 Y(1 - R^2) \cdot (N|N - M)$

< 0 Both sets of estimates are used

5(A)

Columns

Contents

1-3

Number of X values to be read

5(B)

(10F8.0) X values, 10 per card. Use as many cards as necessary.

5(C)

Columns

Contents

1-8

Smallest value of X (min)

9-16

Largest value of X (max)

19-24

N number of steps from X (min) to X (max). Interval size = $\frac{X \text{ max} - X \text{ min}}{N \text{ STEPS}}$

N STEPS

5(D)

For $K \geq 2$, a series of K integers, corresponding to years (i.e., 62 = 1962) are read using the format (2513). A range of years may also be used as input for this option. When the integer appearing in the field of the second year is preceded by a minus (-) punch, the absolute value of this index determines the upper bound of the range. For example, the value 62-70 would include all observations in the range of 1962 through 1970 as input for the analysis.

[Card set 5 and its subsets may be repeated for analysis of additional X, Y pairs of variables included in the original N x M data matrix.]

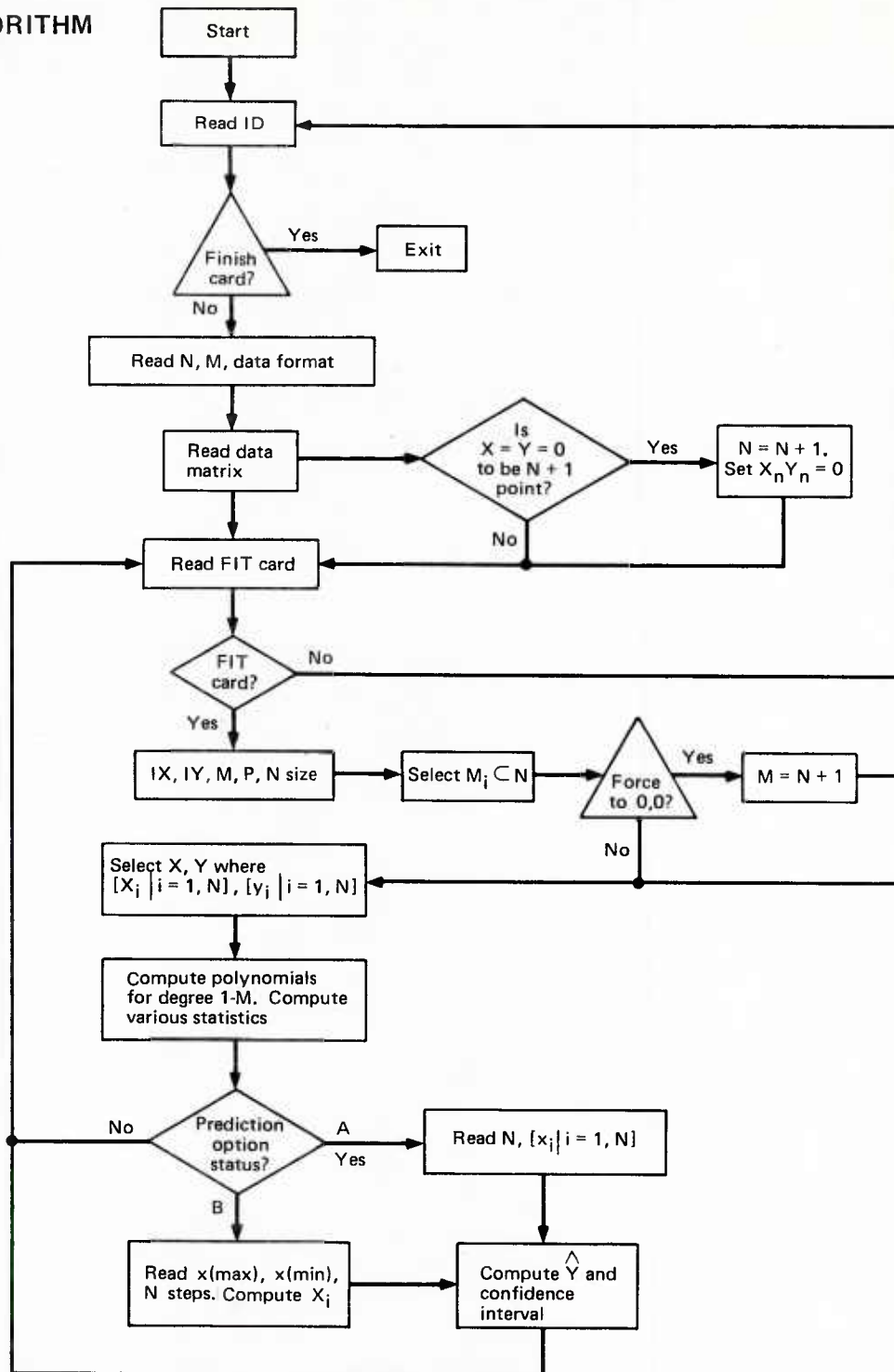
Sequence

6. END FIT card--a blank card to indicate that no more "fits" are required.

[Card sets 1-6 may be repeated for analysis of additional data matrices.]

7. FINISH card--punch FINISH in columns 1-6 to indicate end of run.

ALGORITHM



COMPUTATIONAL PROCEDURE

The following formulas were used to compute the coefficients for the polynomials and to perform various statistical tests on the data.

Let

S_{ij} = the ith sampling unit (FY) for the jth variable (category) where
 $\left[\begin{matrix} i = 1, N \\ \text{elements} \end{matrix} \right]$ and $\left[j = 1, M \right]$ and S is the $N \times M$ matrix of data

x = a vector of N data points constituting the independent variable (predictor), and

y = a corresponding vector containing the N elements of the dependent variable (criterion); such that

$X \in S$ and $y \in S$, with $X \neq Y$.

A polynomial of the form

$$(1) \quad \hat{Y}_i = b_0 + b_1 X_i + b_2 X_i^2 + \dots + b_m X_i^m$$

is generated by least squares regression. That is, M ($M \geq 2$) vectors of powers of X are generated such that the ith vector is equal to X^i .

If we let R_m indicate the matrix of intercorrelations of the m vectors of X , and V_i , a vector of correlations between Y and each X vector, then the standardized regression coefficients, β_i may be obtained from the relationship,

$$(2) \quad \beta_i = R^{-1} V_i$$

This provides (standardized) least squares estimates of the m coefficients of the polynomial shown in (1). Raw score coefficients are then obtained from the relationship,

$$(3) \quad b_i = \frac{\sigma_y \beta_i}{\sigma_{X_i}} \quad \text{where } \sigma_i \text{ represents the standard deviation of the } \underline{i}\text{th}$$

vector of scores.

The intercept, b_0 , is determined by,

$b_0 = M_y - \sum_{i=1}^m (M_{xi} b_i)$, where M_{xi} represents the arithmetic mean of the i th vector of scores.

Other statistics supporting (1) are also obtained.

The squared multiple correlation of Y with $(x_i | i = 1, m)$ is given by

$$R^2_{y \cdot x} = \sum_{i=1}^m \beta_i^2 v_i,$$

which can be interpreted as the proportion of the variance in Y jointly accounted for by the different powers of X.

The F test of (statistical) significance of $R_{y \cdot x}$ is determined by,

$$F = \left(\frac{R^2_{y \cdot x}}{m} \right) \bigg/ \frac{1 - R^2_{y \cdot x}}{n - m - 1}$$

with $(m-1)$ and $(n-m-1)$ degrees of freedom. (Here, $M = N$).

The standard error of estimate (at the mean of x) of $R_{y \cdot x}$ is

$$SE = \sigma_y \sqrt{1 - R^2_{y \cdot x}}$$

Two tests for indicating the gain afforded by adding another term to the polynomial are:

$$\frac{(R^2_{y \cdot x_m} - R^2_{y \cdot x_{m-1}}) (n-m-1)}{1 - R^2_{y \cdot x_m}}, \text{ which is distributed as } F \text{ with } 1$$

and $(n-m-1)$ degrees of freedom; and the

$$\left[\frac{1 - R^2_{y \cdot m}}{1 - R^2_{y \cdot (m-1)}} < 1 - 1 / (n-m-1) \right]$$

"Wherry inequality," which may also be used as a "rule of thumb" for interpreting the gain afforded by using an m vs $m + 1$ degree polynomial (see reference (c)).

Finally, the percent of $R^2_{y \cdot x}$ accounted for by a given x_i is

$$\beta_i \left(\frac{V_i}{m} \right) x \quad 100 .$$

Confidence bounds of \hat{Y} , given an x , are computed as follows:

Let:

$\overline{SE}^2_{y.}$ = Squared standard error at (mean) x based upon m observations.

x = deviation score; $x^k = \bar{x}^k$

$C_{k \times k}$ = weighted variance -- covariance matrix, containing elements:

$$\begin{array}{ccccccc} c_{11} & c_{12} & \cdots & \cdots & \cdots & \cdots & c_{1k} \\ c_{21} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ c_{kl} & \cdots & \cdots & \cdots & \cdots & \cdots & c_{kk}, \end{array}$$

where, $c_{12} = c_{21}$; and,

$$c_{ij} = \frac{1}{n} \sum_{l=1}^n \sum_{l=1}^n x^i_l x^j_l$$

v_k = a vector containing K elements of x

$$\left[x, x^2, \dots, x^k \right]$$

The standard error of an individual forecast, $X.$, represented by $s_{y.f}(x.)$, is given by

$$\left[\overline{SE}_y^2 \left(1 + \frac{\overline{SE}_y^2}{n} + \epsilon \right) \right]^{1/2},$$

where $\epsilon = v_k^T \bar{C}^{-1} k x k v_k$ (Read T as transpose.)

The α -probability confidence limits of \hat{Y} , predicted by the regression of Y on X, are obtained by

$$\hat{Y}_\alpha = \hat{Y} \pm \alpha s_{y.f}(x.)$$

PROGRAM ITERFIT

GENERAL DESCRIPTION

ITERFIT combines the capabilities of least squares (L. S.) curve fitting and profile similarity analysis to determine the effect of adding a (disparate) data point, near the origin ($X = Y = 0$) or elsewhere, to the shape of a curve describing an original X, Y relationship. The "original" curve is described by a polynomial, P_0 , of degree m (here, $m = 2$) determined by a least squares regression for the original N pairs (X, Y) of data points. Another polynomial P_1 , also of degree m , constituting the first iteration, is generated by using $N + 1$ pairs of points. Differences in the shapes of the profiles between the N pairs of data points generated by P_0 and P_1 can then be compared with a zero-order product-moment correlation coefficient, and the elevations, or distances between profiles, with a t test. The routine uses the L.S. regression method, as described in PROGRAM REGFIT. Comparisons among the various iterated profiles are also performed.

Output

- R^2 , F , regression weights (b), Y -intercept, (b_0), standard error of estimate.
- Correlations (as indexes of goodness of fit) between original (N points) and fitted values of Y for each iteration.
- Student's t statistic indicating the significance of the distance between profiles;
- Test of the significance of the differences of iterated profile correlations and elevations.

INPUT REQUIREMENTS

- X and Y data points (variable format)

Limitations

- The array containing original (X, Y) values is limited to 50 variables and 20 observations per variable. (Larger arrays can be accommodated by recoding a number of DIMENSION Statements.)
- $1 \geq m \leq 10$
- Missing observations are treated as real, zero values.

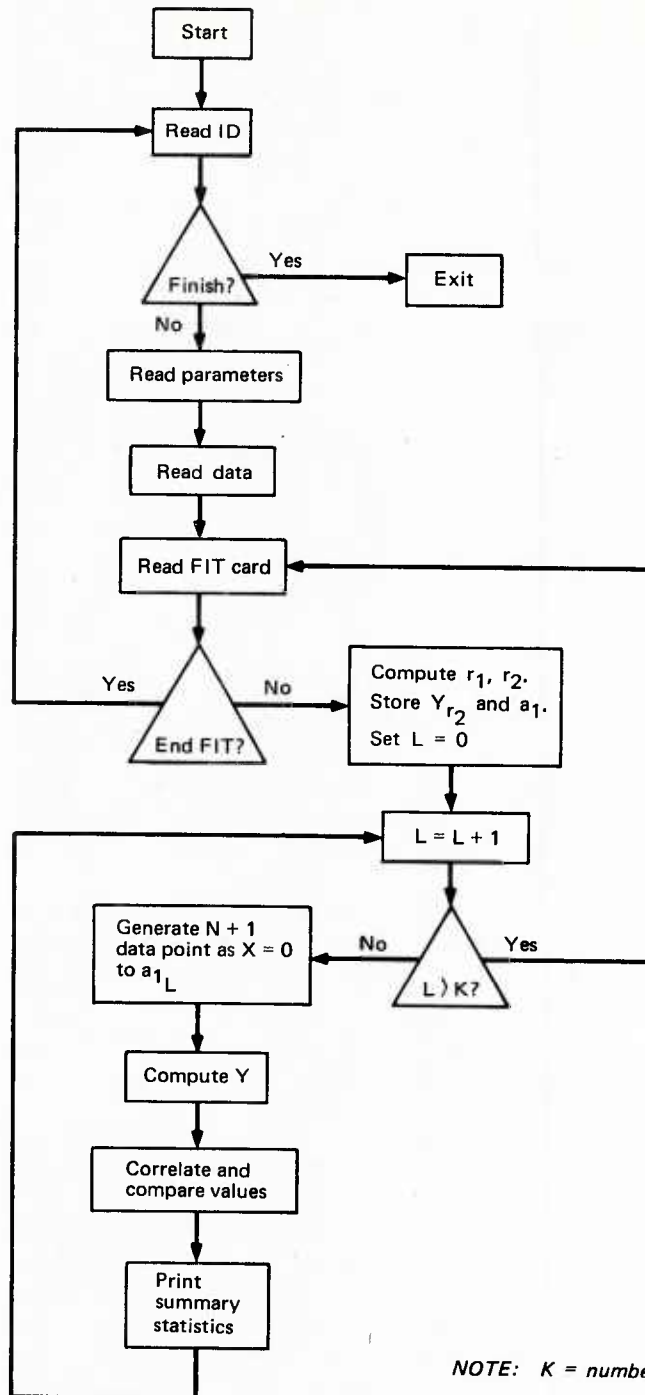
CARD ORDER*

1. Title card
2. Parameter Specification card
3. Variable format card
4. Data (if on cards)
5. FIT card(s)
6. END FIT card
7. FINISH card

SPECIFIC CARD PREPARATION

<u>Sequence</u>	<u>Contents/format</u>
1.	Title card. Any alphanumeric information in columns 1-80.
2.	Parameter specification
<u>Columns</u>	<u>Contents</u>
1-3	Number of observations (n)
4-6	Number of variables
7-9	LUN of input data (blank assumes cards)
10-12	Number of iterations
3.	Variable format card. Use columns 1-80.
4.	DATA (if on cards)
5.	FIT cards
1-6	Any non-blank characters
7-9	Index of X (independent) variable
10-12	Index of Y (dependent) variable
13-15	M, the degree of the highest order polynomial desired
16-17	BLANK
18-20	Number of iterations (if blank, the value on the Parameter Specification Card, columns 10-12 will be the determining value).

ALGORITHM



NOTE: K = number of iterations.

6. END FIT card. A blank card following the last FIT card is used to signal the last sub-analysis of the data (as in 4).
7. FINISH card. Punch FINISH in columns 1-6 to indicate the end of the job.

COMPUTATIONAL PROCEDURE

The regression coefficients and supporting statistics for the least squares polynomials are computed as reported for PROGRAM REGFIT. Additional statistics, comparing the original with the fitted points and fitted points with other fitted data points, are computed as described below.

Preceding the summary of results for a given iteration, the coordinates of the additional ($N + 1$ th) data point are labeled as $X = 0$, $Y = a$, where a is a value between zero and the intercept of the original linear regression equation. The interval size of the distribution of values assigned to a is equal to:

$$\frac{\text{range}(a)}{k}$$

where k is the number of iterations, or steps, requested by the user, and $1 \leq k \leq 50$. The labels appearing on the computer output are as follows.

The label, R-FIT, refers to the product-moment correlation between the points of the original fitted quadratic and the iterated fitted quadratic equations based upon the original N data points.

$R(0)$ refers to the product moment correlation between the original N data points and the fitted data points resulting from a given iteration.

The T-FIT values are Student's t statistics ($N - 1$, d.f.) comparing the mean differences of the values of the fitted quadratics with the original and iterated equations.

The $T(0)$ is analagous to T-FIT with the exception that the values of the fitted quadratic are compared to those of the original N data points.

A matrix comparing the results of the different iterations is also displayed. The values in the diagonal are the R-FIT values described above. The values below the diagonal are the product-moment correlations among the fitted quadratic values for each iteration. The entries above the main diagonal are Student's t statistics for the test of the hypothesis that

$$(r_{i0} = r_{j0}), \text{ where}$$

- i represents the fitted values generated from the i th iteration,
 j represents the fitted values generated from the j th iteration, and
 0 represents the original values of the N data points.
 t is computed by

$$t = \frac{(r_{i0} - r_{j0})}{\left[\frac{(n - 3)(1 + r_{ij})}{2 \left[1 - r_{ij}^2 - r_{i0}^2 - r_{j0}^2 + 2(r_{i0}r_{j0}r_{ij}) \right]} \right]^{\frac{1}{2}}}$$

MANPOWER UTILITY ROUTINE

PROGRAM MANPOWER is a utility routine that organizes the 9 source programs of the MANPOWER Program File. Execution of any program may be accomplished by following the steps outlined below.

1. Instruct the computer operator to mount CNA tape reel number 1834 on logical unit number 14. This input tape is written in BCD with a density of 556 bpi.
2. Include these "systems" cards:

$\frac{7}{9}$ JOB, etc.

$\frac{7}{9}$ EQIP 14 = (SOMAR PROGRAMS), HI

$\frac{7}{9}$ LOADMAIN, 14

3. Precede the control cards for each program of interest with a card containing the name of the program punched in card columns 1 through 8.
4. Following the last control card of the last program, include a card with the characters ENDSOMAR in the first 8 columns.

PROGRAM INTERFAC

GENERAL DESCRIPTION

This utility routine will read and store an array of data and output selected elements from the array. Variable formatting is utilized in both the input and output phases. The routine may be used to reformat a data file, select, sort and/or merge up to a 3-way categorization of the data.

Input/Output

- Equal length, unblocked BCD records.

Limitations

- Uses a single FORTRAN READ Statement to input an array of N elements.
- N is the product of KA, KB, and KC, defined by the user at execution time.
- $N \leq 10^3$

CARD ORDER

Set

1. Parameter specification
2. Input format
3. Data (if on cards)
4. Output selection
5. Output format

SPECIFIC CARD PREPARATION

<u>Sequence</u>	<u>Contents/format</u>
1	Parameter specification
<u>Columns</u>	<u>Contents</u>
1-3	LUN of input
4-6	LUN of output
7-9	Number of elements in first subset of output array
10-12	Number of elements in second subset of output array
13-15	Number of elements in third subset of output array
16-18	Number of elements in first subset if input array

19-21	Number of elements in second subset of input array
22-24	Number of elements in third subset of input array
25-26	Index of first subscript used for ordering output
27-28	Index of second subscript used for ordering output
29-30	Index of third subscript used for ordering output

<u>Sequence</u>	<u>Contents/format</u>
2	Input format--punch FORMAT statement with an open parentheses beginning in column 1
*3	Data (if on cards)
4	Output selection
<u>Columns</u>	<u>Contents</u>
1-3	Index of 1st elements
4-6	Index of 2nd elements (etc., for up to 25 elements/card.)
<u>Set</u>	
a.	$(a_i \mid i = 1, NA)$
b.	$(b_j \mid j = 1, NB)$, and
c.	$(c_k \mid k = 1, NC)$
5	Output format--punch FORMAT statement with an open parentheses beginning in column 1.

ALGORITHM

Data are represented in a 3-dimensional array, $X(I, J, K)$. The range of subscripts are:

$(I = 1, NI)$
 $(J = 1, NJ)$
 $(K = 1, NK)$

Data are read into a linear array Y , which is EQUIVALENCED TO X , by the statement,

READ (LUN, FORMAT) (Y(L) , L = 1, N) .

Here, N is defined as the products of $NI \times NJ \times NK$. The relationship between X and Y is illustrated in table 1.

* See section on indexing procedure for input/output arrangements of data.

TABLE 1

X, Y RELATIONSHIPS FOR NI = 3, NJ = 2, NK = 2

<u>X</u>			<u>Y</u>
(I)	(J)	(K)	(L)
1	1	1	1
1	1	2	2
1	2	1	3
1	2	2	4
2	1	1	5
2	1	2	6
2	2	1	7
2	2	2	8
3	1	1	9
3	1	2	10
3	2	1	11
3	2	2	12

Output Phase

An output hierarchy, determining the ordering of the indexes, I, J, K is provided by the user. The digits 1, 2, 3 are used as codes for the indexes, I, J, K. For example, the configuration of indexes, 2, 1, 3, would produce the following output, for all levels of J, I, and K for the data presented in table 1.

Element X			<u>Order of Output</u>
(I)	(J)	(K)	
1	1	1	1
1	1	2	2
2	1	1	3
2	1	2	4
3	1	1	5
3	1	2	6
1	2	1	7
1	2	2	8
2	2	1	9
2	2	2	10
3	2	1	11
3	2	2	12

Note that the configuration represents a lexicographical ordering of the hierarchy of the range of subscripts.

PROGRAM ONTAPE

GENERAL DESCRIPTION

This program loads cards onto magnetic tape and produces a count of the number of records transferred onto the tape (used as input to process Five Year Defense Program (FYDP) data).

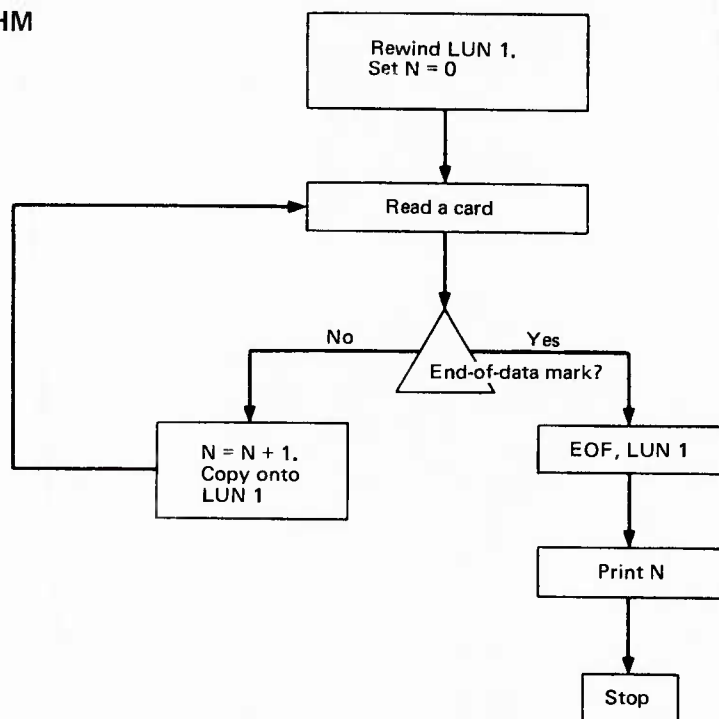
Output

- 80-character, unblocked, BCD, tape records on Logical Unit Number (LUN) 1;
- Card count;
- END-OF-FILE (EOF) after last record copied.

Input

- Data cards to be copied;
- END-OF-DATA card. (Card punched with x in columns 1-8 as a signal that the input file is exhausted.)

ALGORITHM



REFERENCES

- (a) Cattell, Raymond B., Coulter, Malcolm A. and Tsujioka, Bien, "The Taxonometric Recognition of Types and Functional Emergents," in Cattell, Raymond B., editor, Handbook of Multivariate Experimental Psychology, Rand McNally, Chicago, 1966
- (b) Lorr, Maurice and Radhakrishnan, B., "A Comparison of Two Methods of Cluster Analysis," Educational and Psychological Measurement, Vol. 27, pp. 47-53, 1967
- (c) Thorndike, R. L., "Personnel Selection", Wiley, New York, 1949
- (d) INS Study Number 31, "Estimation of Navy Support Manpower Requirements", Confidential (to be published.)

U137702

FOR OFFICIAL USE ONLY

FOR OFFICIAL USE ONLY